Reviewer's report

Title: Cigarette Smoking Prevalence in US Counties: 1996-2012

Version: 1 Date: 11 December 2013

Reviewer: Sean Hu

Reviewer's report:

This study reported findings on estimates of current cigarette smoking prevalence at county level by using validated small area estimation methods to BRFSS data. The findings and small area estimation methods reported were informative and significant from survey research perspective. There are some suggestions to improve the manuscript.

1. There are several missing or typo at several places as follows:
   1.1. At the second line of the section “Bias Correction for Wireless-Only Households”, a right parenthesis is missing after “[22]”.
   1.2. A word “than” is missing for the sentence “the prevalence estimates derived from the 2011 sample were higher than those derived from 2010 sample”.
   1.3. At the second line of the section “Bias Correction for Wireless-Only Households at the Results”, should “landline” be “wireless-only”? 

2. At the section of “Unit of analysis, why the authors chosen respondents aged 20 and over instead of those aged 18 and over? Usually, several national surveys including BRFSS and NHIS always report current smoking prevalence for people aged 18 and over at national and state levels. Choosing respondent aged 20 and over made comparisons more difficult or inappropriate. Can authors briefly address the reason why authors chosen respondents aged 20 and over for the study.

3. At the section of “Model Validation and Performance assessment” in Methods, the authors used “gold standard estimate” for referring to a direct estimate from a pooled 900 respondents. It may mislead readers because this direct estimate may not be a best estimate. So, “direct estimate” may be a more appropriate term.

4. The description of “Bias Correction for Wireless-Only Households” is good. However, the authors pointed out that “We assumed that this bias has increased linearly with time from no bias in the year 2000 to the level measured in 2011 and used this assumption to calculate corrected estimates in each year from 2001 to 2010”. However, how this bias linearly increased across year is not clear? Since this increase is only assumption, it may limit the data findings.

5. In the Methods section, there is no description of age-standardizing the estimates instead of one sentence at the appendix. It is better to briefly describe it at the Methods section.

6. Comments on methods:
This paper applied a series of county-level binomial regression models to estimate the county-level prevalence of smoking outcomes. By nature, these small area models are (area-level) ecological regression models. There are at least four major (6.1-6.4) and three minor concerns (6.5-6.7) for this ecological approach in this paper.

6.1. First, the county-level small area models ignore the uncertainties associated with sample sizes in model estimations. A larger BRFSS sample size for a specific county demographic group, should be weighted more in the final model estimation than a smaller one. However, current approach would overweight the small sample size county demographic groups and underweight large sample size county demographic groups. The ignored “weights” of sample sizes could result in significant potential bias in model parameters.

6.2. Second, the county-level small area models strongly assume that all the county-level contextual variables (covariates) have universal impact on smoking outcomes across the whole US and thus could suffer substantial ecological bias in model prediction. Thus, the validation of county-level model-based estimates is critical and necessary. Although the BRFSS was not designed for county-level estimates, the county-level model-based estimates of smoking outcomes, when aggregated to state level, should be consistent with the state-level direct survey estimates from BRFSS. BRFSS were designed for reliable and accurate state-level estimates. If there are significant differences between model-based and direct survey estimates of smoking outcomes at state-level, the county-level small area models might be misspecified. The model validation implemented in this paper did not fully meet this goal.

6.3. Third, National Cancer Institute (NCI) combined BRFSS and National Health Interview Survey (NHIS) to generate county level small area estimates of current smoking outcomes by sex for the time periods 1997-1999 and 2000-2003 (http://sae.cancer.gov/estimates/current.html)[1]. It could be a good data source for an external validation of county-level model-based small area estimates of current smoking in this paper. The comparison between NCI estimates and the estimates from this paper may be able to evaluate the bias from the missing wireless-only population and even the population without phone in BRFSS.

6.4. Fourth, the county-level small area models included only one state variable (state-level cigarette sales per capital) to account for state-level contextual influence. As we know, there are substantial variations across states in smoking outcomes. The single state-level variable may not be able to account for state-level variations (state-level contextual effects) after controlling county-level contextual influence. If state-level variance component is still significant in the final models, a state-level random effect should be included in the models for final prediction.

6.5. The author might check the correlations between all the county-level variables to avoid the potential multicollinearity in the final models.

6.6. The aggregation of multiple years of BRFSS data for small area estimation could result in additional bias when the outcomes of interest have complex temporal trends. Single year BRFSS could produce reliable county-level small
area estimates via a unit-level multilevel logistic small area models [2, 3]. It could avoid additional linear assumption about temporal trends in smoking outcomes in this paper.

6.7. Total cigarette smoking and current total cigarette smoking are different smoking outcomes and should be consistent.

References

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests.